ID23- AUTOMATIC FISH COUNTING FROM UNDERWATER VIDEO IMAGES: PERFORMANCE ESTIMATION AND EVALUATION

MARINI S.[125], AZZURRO E.[126], COCO S.[43], DEL RIO J.[182], ENGUÍDANOS S.[88], FANELLI E.[56], NOGUERAS M.[222], SBRAGAGLIA V.[107], TOMA D.[181], AGUZZI J.[99]

Abstract – Cabled observatories offer new opportunities to monitor species abundances at frequencies and durations never attained before. When nodes bear cameras, these may be transformed into the first sensor capable of quantifying biological activities at individual, populational, species, and community levels, if automation image processing can be sufficiently implemented. Here, we developed a binary classifier for the fish automated recognition based on Genetic Programming tested on the images provided by OBSEA EMSO testing site platform located at 20 m of depth off Vilanova i la Gertrú (Spain). The performance evaluation of the automatic classifier resulted in a 78% of accuracy compared with the manual counting. Considering the huge dimension of data provided by cabled observatories and the difficulty of manual processing, we consider this result highly promising also in view of future implementation of the methodology to increase the accuracy.

Keywords –Cabled observatories; manual fish counts; automated fish counts; pattern recognition

I. INTRODUCTION

Persistent climatic or human-induced environmental changes can produce long-lasting modifications in species behavior, with pervasive effects on population distributions and abundances [1]. These effects can be particularly apparent in fish, due to their high mobility, which can allow entire populations to respond rapidly to environmental changes [2].

Traking these changes at the fine temporal scale is today a need of community studies, with clear implications for fisheries and, in a broader sense, for ecosystem management. Yet, the cabled seafloor video-observatory technology offers new opportunities to monitor species abundances at frequencies and durations never attained before [3]. The largest existing networks, are to date the European Multidisciplinary Seafloor and water column Observations unit (EMSO; www.emso-eu.org), the Ocean Network Canada unit (ONC; www.oceannetworks.ca/), and the Japanese Dense Oceanfloor Network System for Earthquakes and Tsunamis (DONET; http://www.jamstec.go.jp/donet/e/). When nodes bear cameras, video-counted individuals can be used as indicator of population rhythms and then related to surrounding habitat conditioning in terms of cause-effect principles, by measuring at the same time different oceanographic, chemical, and geologic parameters.

Video cameras may be transformed into the first sensor capable of quantifying biological activities at individual, opulational, species, and community levels, if automation image processing can be sufficiently implemented [4,5]. A consolidated scientific literature exist in computer vision and pattern recognition [6,7] and rapid progresses have been made in implementing automatic counting and classification . These technological advances, combined with the continuous improvement of the hardware performances provide a valid support for investigating the big amount of data provided by the cabled observatories and for understanding the complex dynamics of the underwater ecosystems;

Actually, within the Pattern Recognition literature, very few journal papers propose methodologies for recognizing fishes, and all of them are based on small datasets where the images of the fishes are really easy to be processed (e.g. no turbidity, no fouling, shape and texture of the fishes are clearly visible, etc.). Conversely, there are abundant conference-papers proposing different Martech 2016. recognition approaches, but also with very simple data sets (few images processed, few days analysed, single species approach, etc.).

Here, we studied and developed a binary classifier for the fish automated recognition based on a Genetic Programming supervised machine learning approach. The classifier has been trained on the images acquired by the OBSEA EMSO testing site platform (www.obsea.es) located at 20 m of depth off Vilanova i la Gertrú (Spain).

The high variability of environmental conditions in coastal environments, especially related to turbidity, hydrodynamics, light and fouling, makes it challenging the study of the habitats and actually automatic counting has been never tested in these variable environments. Moreover the Marine Strategy Framework directive (MSFD, 2008/56/EC) through the technical guidance for monitoring (JCR 2014, Report EUR 26499 EN) identified in high-definition cameras promising approaches for biodiversity monitoring (Descriptor 1), which could have implications also for other descriptors, such as 2 (alien species), 3 (commercially exploited fish and shellfish) and 10 (marine litter). Thus the implementation of an automation protocol could have several important consequences for monitoring programs at a European or ider scale. Within this context, the aims of this work are to:1) estimate and compare the performance of the method here tested by contrasting automatic vs. manually-recorded data; 2) to assess the performance of this methods in different environmental conditions/constrains (i.e., clean vs. turbid water; clean vs. dirty camera, daytime vs. nighttime)

II. MATERIALS AND METHODS

The Observatory of the Sea (OBSEA) is represents one of the few coastal multiparametric observatories currently active. OBSEA was launched in 2009 off the Catalan Coast in Spain (western Mediterranean: 41°10'54.87" N and 1°45'8.43" E) and deployed at a depth of 20 m within a marine reserve (Colls Miralpeix Marine Reserve) for details on the platform, instruments and sensors see [4].

A time-series consisting of about 12,900 images, collected every 30 minutes by the OBSEA observatory in the 2012, has been visually inspected by expert biologists in order to estimate the number of fishes captured by the images. The camera was oriented in front of an artificial reef, omposed by concrete blocks. Additionally, the formation of fouling on the camera protecting dome preventing from the acquisition of good images.

The data resulted from the visual inspection of the OBSEA time-series has been used as training and validation set for learning an automatic image classifier capable to automatically recognise the fishes present in the images;

Every image has been processed in order to extract the image-features used by the automatic image classifier to discriminate fishes from other floating objects and from the seabed and from the fouling.

In order to extract the image-features, a pre-processing step is needed to identify the the image Region of Interest (RoI) potentially containing fishes. Histogram adjustment based on the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [8] is performed, for improving contrast between background and foreground items. The foreground is then segmented from the background using an adaptive thresholding. The foreground binary map is post-processed by opening/closing morphology operators to remove small dots and fill small gaps. Edge detection is performed with a filtering process based on the Sobel operator [9]. The segmentation process is tuned in order to maximize the probability to identify the RoIs potentially containing fishes, allowing for an unavoidable incidence of false alarms.

The image-features extracted form the RoIs belong to two groups: geometrical, based on the shape of the blob; textural, based on the grey levels distribution inside (and outside) the RoI. The geometrical features are: length of the minor semiaxis (sAxm), bounding box minor dimension (axm), bounding box major dimension (axM), eccentricity related to the semiaxis ratio (ecc). Other geometrical features, are the RoI solidity defined as the area ratio between the RoI and its convex hull (sol), area (areap), perimeter (per), radius histogram shape index defined as the ratio between the standard deviation and the mean value of the boundary (hstI), enthropy (ent). The textural features, are: exterior-interior contrast defined as absolute difference between the averaged grey levels inside/outside the RoI (ctrs), grays level standard deviation (stdg1), contrast index defined as the ratio between standard deviation and mean of the grey levels (stdg), gray levels enthropy (entg).

The machine learning methodology used in this work is based on Genetic Programming (GP).

GP is an evolutionary computation methodology capable of learning how to accomplish a given task. GP generates the task solutions starting from an initial population of randomly generated func- tions, based on a set of mathematical primitives, constants and variables. The initial solutions are improved by miming the selection processes that occur naturally in biological systems through the Selection, Crossover and Mutation genetic operators [10]. In the proposed work, the set of mathematical operators S is used to generate binary classifiers expressed as mathematical functions, whose variables correspond to the image-features discussed above. Details on this method can be found in [11]

The training experiments have been performed within a Kfold cross-validation (CV) framework [12] in order to estimate the generalization performance of the automatic image classifier.

III. RESULTS AND DISCUSSION

The image classifier has been trained on 700 randomly sampled images, corresponding to the 5.4% of all the images acquired in 2012. The reamaining 11600 images have been used for testing the learnt classifier.

The training and validation performance of the image classifier have been estimated by computing the average and standard deviation of Accuracy:
ACC = (T P +T N)/ (T P +F N +F P +T N), the True
Positive Rate TPR = TP / (T P +F N) and False Positive
Rate FPR = FP / (F P +T N), where TP, FP and TN
represent True Positive, False Positive and True Negative
recognitions respectively (Table 1).

	Mean	SD
ACC	0.78	0.02
TPR	0.62	0.08
FPR	0.13	0.04

Table 1. Mean and standard deviations (SD) of performance indicators. ACC=accuracy; TPR= true positive rate; FPR= false positive rate.

If we train the classifier to capture fishes that vanish into the background, we increase the risk to mistake the fouling as fishes (false positives), while if the camera is clean some false positives occur, without affecting the performance. Generally the classifier was effective in fish identification in reduced fouling conditions (Figure 1).

Fig 1. An example of image captured with the camera without fouling

If the scene is crowded the classifier identifies group of fishes as they were just one fish (Figure 2). That is why in the time series corresponding to crowded periods, the number of recognized fishes is smaller than the number of observed fishes. Nevertheless the trend of the fishes is still coherent.

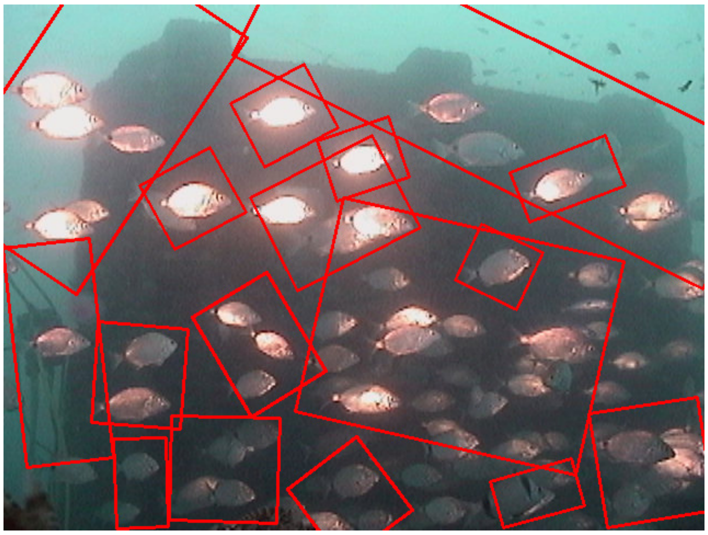

Fig 2. An example of an image inaccurate identification by the classifier of a group of fish as one fish.

Overall, the performance evaluation of the automatic classifier resulted in a 78% of accuracy compared with the manual counting (Figure 3).

IV. CONCLUSIONS

The OBSEA time-series are very relevant because only few observatories in the Mediterranean Sea provide so much information (AcquaAlta, http://www.ismar.cnr.it/infrastructures/piattaforma-acquaalta/) and because the acquisition conditions are highly variable and make these time-series a real dataset and not a toy example, as in similar works using this methodology.

The automatic recognition of fishes provides good results especially when fouling on the camera is limited, while improvements are needed in order to reduce the false positives caused by fouling and for managing large school of fishes that make the scene crowded. The classifier learnt from the images acquired in 2012 will be also validated on the images acquired from 2013 to 2016. Finally, the methodology proposed in this work for learning the automatic classifier will be also applied on the image timeseries acquired by other observatories (e.g. the Acqua Alta platform).

V. REFERENCES

[1] Peer A.C., Miller T.J. Climate Change, Migration Phenology, and Fisheries Management Interact with Unanticipated Consequences. North American Journal of Fisheries Management vol. 34, pp. 94– 110, 2014.

[2] Cheung W. W. L., Sarmiento J. L., Dunne J., Frolicher T. L., Lam V. W. Y., Deng Palomares M. L., Watson R., et al. Shrinking of fishes exacerbates impacts of global ocean changes on marine ecosystems. Nature Climate Change vol.3, pp.254-258, 2013.

[3] Aguzzi, J. et al. Challenges to assessment of benthic populations and biodiversity as a result of rhythmic behaviour: video solutions from cabled observatories. Oceanography and Marine Biology: An Annual Review (OMBAR) 50, 235-286 (2012).

[4] Aguzzi, J. et al. Coastal observatories for monitoring of fish behaviour and their responses to environmental changes. Review Fish Biol. Fisheries: In Press doi: 10.1007/s11160-015-9387-9 (2015).

[5] MacLeod N., Mark Benfield, Phil Culverhouse, "Time to automate identification", Nature 467, 154–155, 2010.

[6] Duncan K., S. Sarkar, Saliency in images and video: a brief survey, Computer Vision, IET 6 (6) (2012) 514–523.

[7] Huang Y., Z. Wu, L. Wang, T. Tan, Feature coding in image classification: A comprehensive study, Pattern Analysis and Machine Intelligence, IEEE Transactions 36 (3) (2014) 493–506.

[8] Reza, A.M. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology 38(1), 35–44 (2004)

9] Walther, D., Edgington, D.R., Koch, C.: Detection and tracking of objects in underwater video. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Pro-

ceedings of the 2004 IEEE Computer Society Conference on. vol. 1, pp. I–544. IEEE (2004)
[10] Koza, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA (1992)
[11] Corgnati L., L. Mazzei, S. Marini, S. Aliani, A. Conversi, A. Griffa, B. Isoppo, E. Ottaviani, Automated gelatinous zooplankton acquisition and recognition, in: Computer Vision for Analysis of Underwater Imagery (CVAUI), 2014 ICPR Workshop on, IEEE, 2014, pp. 1–8.
[12] Kohavi R, editor A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International joint conference on Artificial intelligence 1995.

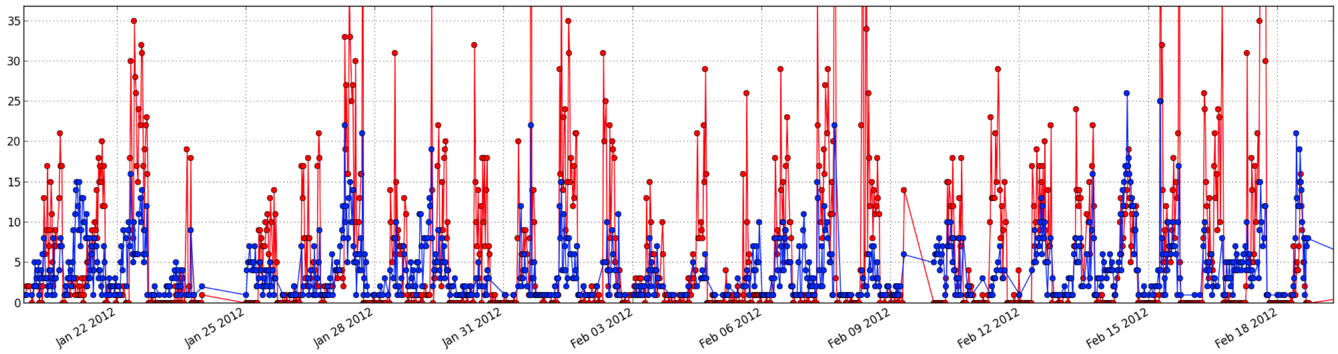

Fig 4. a) Bathymetry generated from the multibeam data gathered by Sparus II data while navigating